Work Project, presented as part of the requirements for the award of a Master's Degree in Management from the Nova School of Business and Economics and Insper.

DETERMINANT FACTORS OF EVASION IN BRAZILIAN'S PRIVATE HIGHER EDUCATION INSTITUTES

HENRIQUE ALVARES COSTA

Work project carried out under the supervision of:

Prof. Carlos Santos (Nova SBE)

Profª. Adriana Bruscato (Insper)

22-05-2020

Abstract

This paper studies some determinants of evasion in Brazilian's Tertiary Education. It addresses the impact that being a freshman, studying in Distance Learning, enrolling in STEM (Science, Technology, Engineer and Mathematics) subjects and having financial aid have on drop-out rates, focusing in the Private Sector. With a database of 1.662.846 datapoints and 13 variables (including our dependent), and using Logistic Regression, I concluded that Distance Learning, being a freshman and access to financing significantly impact evasion rates (where the first raises drop-out chances, while the others decreases them). Although STEM also influences evasion, it doesn't have a strong impact.

Keywords

Evasion, Logistic Regression, Brazil, Tertiary Education, Private Sector, Attrition, Drop-out.

This work used data available through IBGE (Brazilian Institute of Geography and Statistics), which contains official statistical information regarding education in Brazil.

1. INTRODUCTION

According to the Higher Education Census (2014), there was a significant increase in the number of Higher Education Institutions (HEIs) in Brazil, from 882 in 1980 to 2368 in 2014, representing an increase of 268.5%. Of these, 87.4% are made up of Private Institutions, and the other 12.6% are related to Public Institutions. Still demonstrated by the Higher Education Census (2013), there was also an increase in the number of courses offered by HEIs, both in the classroom and in distance learning during the period from 2008 to 2014 (27% and 112%, respectively).

Thus, it is noted that most of the institutions that compose the framework of HEIs in Brazil are from the private sector and, as Campos (2018) points out, it is evident that they compete for students and financial resources, with the goal of increasing quality and variety of services offered, and adapting the guidelines according to current needs to ensure competitive advantage.

However, accompanied by this higher education's set up in the country, there is still a huge problem when it comes to tertiary education, especially in the private sector: the number of dropouts. According to the Higher Education Census (2018), provided by IBGE (Brazilian Institute of Geography and Statistics), of 7.910.755 students enrolled in Private Higher Education's Institutions in Brazil in 2018, 2.371.940 students dropped out, representing a loss of around 30% (almost a third) of all the students. Taking into consideration that, as Campos, Henriques and Yanaze (2017) rightly points out, a country's educational system is, as a rule, consistent with that country's governmental system, this can be seen as a serious problem. In addition, future well-being, public policies and national purpose are closely related to university's formations (Selden, 1965). Finally, education is one of the main pillars for the economic, social and political development of any nation (Shrivastava, 2014).

As can be seen, evasion is an extremely serious problem that must be tackled; this view is in line with the idea defended by Beer and Lawson (2017), that the evasion of higher education students remains a costly and significant challenge for institutions around the world. In addition, if universities are spending resources on projects and plans aimed at improving student retention with limited success, they face a threefold problem: loss of income from evading students, futile investments in organizational funds and reputational damage.

Still according to Beer and Lawson (2018), previous researches realized two areas that needed investigation related to student's dropout. The first one considers that evasion can be conceptualized as a "perverse problem"; this means that it is a problem considered complex and multifaceted, and that cannot be solved using traditional problem solving approaches, as they are problems outside of Universities' and Governments' control (personal reasons generally related to professional life or responsibilities; and reasons that appear during the higher education programs themselves, such as learning style and career fit). The second area is composed of factors that are under the control of institutions (which in turn allows action), and is related to how universities can deal with these problems related to evasion. Important to note that this work will focus on the latter, which could be under institutional control.

One of the objectives of this work is to bring these relationships to the educational context of Private Higher Education Institutions in Brazil, with the purpose of analyzing whether the problems and challenges that countries around the world have can be applied to the Brazilian's context. The other goal is to identify and outline possible action plans (which are the responsibility of the HEIs) which could help leverage retention indicators and, consequently, contribute to the country's development, at the same time that it also aids Private Universities to increase their profits.

Therefore, the work is structured as follows: the first chapter was presented in this session, in chapter two the literature review is presented, chapter three presents the database,

treatments performed and methodology that will be used for this article; in the fourth chapter, data and the results of the Logistic Regressions will be analyzed and finally, in the last section, the conclusions and possible managerial actions to leverage the retention indicators will be presented.

2. LITERATURE REVIEW

There is a lot of concern not only regarding access to education, but also a student's success (NAO, 2002). This concern was already emerging in the 1970s, gained strength in the 1990s and continues until nowadays. Considering that students who are currently in tertiary education may eventually become teachers and researchers of other universities a few years from now (Pritchard, 1995), this may be a bad sign for the future of higher education in the world.

Since the 1970s, there were studies on school dropout in Higher Education. At the time, an Institutional Evasion Model (Tinto, 1975, 1987, 1993) was created, offering a theoretical basis to explain the reasons that lead students to evade. This model also influenced the focus on integration and preceded a large body of research on student retention (Bers & Smith, 1991).

According to Tinto's model, demographic and social characteristics of individuals (such as family, individual attributes, prior education to higher education, ...) interact with each other and influence personal attributes related to student engagement and commitment to its institutions. These attributes and characteristics impact a student's intellectual development and academic performance, which in turn culminates in academic integration. Basically, this model argues that students' perceptions regarding their academic and personal integration are extremely significant influences for decision making as to whether they will persist or evade in higher education. Therefore, in its last works, Tinto (2000, 2002 and 2003) started to defend the idea of developing teaching and academic communities in order to improve the process of

academic and social integration, and this has become a very dominant theme in the literature related to academic retention.

After the first work by Tinto, there was a significant increase in research that supports his hypothesis that the interaction between social and academic factors of individuals has a relevant importance in their decisions regarding Higher Education (see Cabrera et al , 1992; Peterson & delMas, 1996). In addition, many studies based on this model have provided significant empirical evidence, validating the importance of this concept of interaction between demographic and institutional factors for the retention of students in private HEIs (see Pascarella & Chapman, 1983; Pascarella & Terenzini, 1980, 1983).

Despite gaining more strength since 2000, in the 80s and 90s there were already several studies from many countries in order to analyze retention in higher education, where we can mention: United Kingdom (see DfEE, 1998), Australia (see Martin, Maclachlan & Karmel, 2001), Canada, Japan, Germany, France, Italy (see NCES, 1996) and the United States (see Schwartz, 1996). That is why, according to Peterson et al. (1997), from this decade onwards, research on student retention in Higher Education began to proliferate quickly, and this problem began to become international in scope. In addition, this increase in research, according to Grosset (1991), was related both to economic concerns (Student Recruitment and Enrollment Maintenance) and to humanistic and social concerns, since capturing and retaining higher education's students were (and continue to be) very important for the economic well-being of Academic Institutions and countries. Therefore, evasion should be seen as a problem of institutional efficiency, leading both to the creation of strategies with the purpose of remedying it and the need to better understand what reasons (characteristics of students and institutions) contribute to its occurrence.

More recent data shows that the retention problem is still significant in some parts of the world, such as England (see NAO, 2007), Germany (see Heublein, Schmelzer and Sommer,

2008), USA (see Astin and Oseguera, 2002 and American College Testing Program, 1998), France (see Ministry for Higher Education and Research, 2016), Spain (see Ministerio de Educación y Ciencia, 2001), among others.

There is much evidence about many factors that may impact on tertiary education's evasion. However, sometimes it fails to effectively prove the reasons that lead students to a drop out. According to McInnes, James & Hartley (2000), student withdrawal is a complex and often individualized process involving a set of institutional, social and personal factors that, when combined and in constant interaction, lead students to evade.

These factors are wide and they range from individual and personal characteristics (socioeconomic, if you are the first person in the family to attend higher education, distance from the University, if you have children, if you work and study at the same time, age, gender ...) to Universities' characteristics, where it is possible to have a certain level of control, that according to Govendir et al. (2009), can be: quality of relationships between students, supervisors and teachers; students' involvement in extracurricular or socializing activities; communication with other students; students' concerns about future professionals, and the interaction between students and their peers, among others. Complementing this point of view, Roberts (2018) argues that success factors for student retention are not only restricted to excellent learning and teaching, but also to a variety of students' supporting mechanisms focused on putting their experiences ahead of all institutional activities. It is also necessary to understand that different stages of the student's study cycle require different needs and approaches.

In the Brazilian context, discussions about dropout in Higher Education gained significant strength in recent years. As of 1996, greater production on the topic began to be identified, mainly due to the publication of the report of the Special Commission on Studies on Evasion (Mec et al., 1996). However, studies on this topic are scarce when compared to the

USA and Europe and, in addition, most of them analyze retention in the context of Specific Universities (see Fregoneis, 2002; Andriola et al., 2006; Adachi, 2009; Lima Jr., Silveira and Ostermann, 2002) or for specific courses (see Silva et al., 2012; Vanz et al., 2016; Hoed, 2016). The main factors referring to dropout in the Brazilian context, according to these authors, relate mainly to failure in the disciplines, low knowledge about the course, low compatibility between academic calendar and professional activity, family aspects, curricular inadequacy, Science, Technology, Engineering and Mathematics courses, evening courses, socio-economic conditions, among others. It is important to note that these factors are in line with what the international literature defends.

Despite the literature on Higher Education Retention being vast and having many factors that could be analyzed, this study will focus on four main variables: evasion in the first year, evasion in distance education courses, evasion in the area of knowledge composed of Science, Technology, Engineering and Mathematics, and impact of students' funding on tertiary's education dropout. Below, there are more details about each one that will be taken into consideration.

Evasion in the first year

In a study made by Venegas-Muggli (2019), he found that 28% of students in his sample dropped out of their Universities during their first year. Complementing this point of view, most of the retention programs focus especially on first-year students, since it is when, according to Barefoot (2004), a big part of all drop-outs happen. Also, according to Bodin and Orange (2018), a third of students entering the University in France disappeared by the end of the first year. Lassible and Gómez (2008) identified, through a study in Spain, that by the end of the first year of Higher Education students have a 26% greater chance of dropping-out, while in the second and third year, respectively, these rates fall 10% and 6%. After three years, it drops

further to 5%. In South Africa, for example, the dropout rate for first-year students increased from 23.6% in 2000 to 25.6% in 2002 (Department of Higher Education and Training 2016).

This problem was already being highlighted by Kember (1995), arguing that, in one course, the dropout rate is much higher in the beginning of the first year than in the following years, reducing significantly in the next years. Therefore, it is understood why most studies regarding students' experience in tertiary education focus on the first year of graduation (Krause et al, 2005). This is because, as other studies have raised (DEST Strategic Analysis and Evaluation Group, 2004; Wintre et al., 2006), approximately half of dropouts in higher education occur in the first year of studies. This point of view is highlighted and corroborated by several authors (see Spady, 1970; Tinto, 1975; Tinto and Goodsell, 1993; Woosley, 2003). Thus, the first hypothesis of this work is proposed:

H_1: Students in the first year of higher education are more likely to drop out than students in the following years.

Evasion in Distance Learning Courses

According to Joshua et al (2019), online courses are seen as having a significant transformation potential for higher education, decreasing delivery costs and increasing access to the Third Degree for students with social disadvantages. The US federal government estimated that 27% of all university students were taking at least one online course in 2013 (most recent date in the North American Database). In addition, from 2002 to 2012, the number of online Bachelor's degree courses grew from 4.000 (Four Thousand) to 75.000 (Seventy-Five Thousand) in the United States, which corresponded to 5% of all Bachelor's degrees completed in 2012 in the country. (Deming et al. 2015).

Another extremely important data raised by Joshua et al. (2019) is that, in 2016, the University of Illinois started enrolling students in its Online MBA program; Yale University is

also currently developing an online version of its Master of Science in Medical Science course for Medical Assistants. Furthermore, in 2016, more than a dozen excellent North American Universities joined the edX Consortium, initiated by Harvard and the Massachusetts Institute of Technology (MIT) by announcing plans to offer small Masters' Courses online. According to He et al. (2014), more than 6.7 million students of Higher Education are taking at least one online course, and 69% of American Institutions include Distance Education as a crucial part of their long-term strategies.

In addition, as previously mentioned, the Higher Education Census (2014) showed an increase of 112% in courses offered online in Brazil during the years from 2008 to 2014, which indicates the latent potential of Distance Learning courses. In the country's context, this increase is being a decisive factor for an educational revolution, since they provide access to Higher Education to those who do not have many resources, contributing to an improvement in its formation in Brazil.

According to Costa and Cochi (2014), Distance Education in the Brazilian context is part of a social policy that, despite making economic development possible through greater access to the labor market and qualified labor, it reduces inequalities in income, creates opportunities and raises living standards. Based on the studies by Litto & Formiga (2008), the main advantages of this kind of education are: mass quality teaching; efficiency of distance education when compared to classroom education; lower costs; avoidance of larger concentration of students and staff; make access to education in rural areas possible; reduce migration; guarantee quality education where there are no Physical Institutions capable of doing so. Considering that Brazil is a country with large territorial dimensions, with many rural areas and enormous economic inequality, with a deficit of qualified labor, and where Public and Regional Education is sometimes scrapped, the Distance Education's modality has been gaining strength and contributing to the country's development.

However, even with this positive scenario regarding distance education, a dilemma began to emerge regarding this type of teaching and its retention rates. According to Carr (2000), there is a lot of evidence and studies suggesting that the retention and completion rate of distance learning courses are lower than in face-to-face courses, which can be attributed either to demographic issues or to the fact that online courses cannot provide the personal interaction sought by many students.

This dilemma was already raised by Losty and Broderson (1980), who argued about how students in the online modality at higher level around the world were characterized as having a higher dropout rate than students on Campus. In addition, Chyung (2001) has already warned about the fact that high dropout rate in online courses enrollment has been a serious problem in distance education. Complementing this point of view, Swail, Redd and Perna (2003) found that dropout rates in traditional schools remained at 50% for approximately one year, but when taking distance courses into account, this rate increased to 70 % (Sheppard, 2002). Thus, the second hypothesis of this study is proposed:

H_2: Students enrolled in Distance Learning in Higher Education are more likely to drop out than face-to-face students.

Evasion in the knowledge area of Science, Technology, Engineering and Mathematics (STEM)

Another relevant problem highlighted by the literature concerns evasion by area of knowledge. According to Heublein (2014), from 2008 to 2009, 28% of first year students dropped out of the university; of all of them, 31% to 39% (above the average per knowledge area) were in the Science, Technology, Engineering and Mathematics courses. According to the Organization for Economic Cooperation and Development (2008), in many countries around the world this area of knowledge is among those with the highest rates school dropout rate. This represents a significant problem, since especially in western countries, there is a belief that there

is a necessity to increase the number of graduates in this field. This, according to Holmegaard and Dyrberg (2019), raises concerns about a future deficit of graduates in this area of knowledge.

In Germany, for example, in a study regarding the failure of students to complete higher education courses entering the science area between 1999-2001, it was noticed that 28% did not complete their undergraduate course, with some differences between the disciplines. Physics, Sciences, Computer Science, Mathematics and Chemistry lost between 31% and 36%, while Pharmacy, Biology and Geography lost between 6% and 15% of students. Engineering had an evasion rate of 25%, ranging from 16% to 34% depending on the discipline (Heublein, Schmelzer and Sommer, 2008). These significant dropout rates have two problems in particular: they significantly reduce the production of graduated students in the field of Science and Technology, and discourage students from enrolling in courses in those areas, which can be known as "Expectation of Success": knowing that the success rate of such courses is very low, potential newer candidates tend to choose other areas of study (The Organization for Economic Co-operation and Development, 2008). Faced with this scenario, the third research hypothesis emerges:

H_3: Students studying Science, Technology, Engineering and Mathematics are more likely to drop out than students in other areas of knowledge.

Impact of students' financing on school dropout

According to DesJardins et al. (2006) and Dynarski (2002), having access to forms of financial assistance has a positive impact on recruitment and a negative impact on student dropout, especially for low-income students. According to Melguizo (2010), one of the main factors related to the high dropout rates in Colombia was due to the lack of student funding. Cerdán-Infantes and Blom (2007) showed that, according to a study by the World Bank, dropout

rates in the country were 30% lower for financing beneficiaries when compared to non-beneficiaries who shared the same characteristics. Consequently, they concluded that this financing (especially in forms of loan) offered an efficient alternative to increase the enrollment of students, while reducing the number of dropouts from low-income students.

For Saccaro, França and Jacinto (2019), policies created for granting benefits to students in social vulnerability (such as student financing) are very important to increase retention rates in Higher Education. This can be corroborated by the study pursued by Nora (1990), who concluded that the main reason why students drop out of Higher Education in Texas was financial difficulties. Santelices (2016) also proved that there is a positive impact of financial assistance on retention in Higher Education. Therefore, the last hypothesis to be analyzed in this study is proposed:

H_4: Students with government assistance through funding are less likely to escape than students without any help.

Control Variables

Recently, there has been a change in the profile of students attending higher education. Adults are becoming the main students enrolled in this type of institution. According to data from the United States Department of Education, in 1995 more than 25% of undergraduate students were 30 years of age or older. Between 1980 and 1990, participation at this level of education for students under 25 years old increased by 3%, while the participation of students over 25 years old increased by 34%. Another important thing to consider is that part-time enrollments increased by 28% between 1984 and 1994, while full-time enrollments increased by only 14% (NCES, 1995). This data shows that it is very important for Higher Education Institutions to start thinking about how to increase the persistence of students who are older than before, both in the United States and in other countries.

This concern about an increase in the average age of higher education students gains even more strength since many studies show that the older the students are, the more likely they are to evade. Students who are over 21, for example, have a greater chance of dropping out of higher education than those under 21 (Arulampalam, Naylor, and Smith, 2004; NAO, 2007). This point is also highlighted by Jeffreys (2007), where he shows that voluntary evasion is the most common for older students, and that the average age of people who evade is 33 years, while the average of those who are retained is 28 years old. Still, according to Venuleo and Salvatore (2016), in a study about the probability of dropout in a psychology course at a University of Italy obtained through a Logistic Regression, it was noted that the chances of retention decreases as age increases. Lassible and Gómez (2008) also contribute to this point of view, demonstrating that the older the students are, the greater the chances of dropout in Spanish Universities.

Regarding to gender, a study of higher education by the National Audit Office (2007) showed that women have a higher retention rate in higher education than men. Corroborating with this point of view, Arulampalam, Naylor and Smith (2004) found that from a sample of 51.810 medical students, men had an 8% higher chance of giving up than women. From the same point of view, a study by Lassible and Gómez (2008) shows that male students at technical universities in Spain are more likely to drop out of the course than women: the probability of women dropping out is around 80% of the likelihood of men. This can also be proven historically: in New Zealand, during the period 1999-2007, around 36% of students who started their courses in 1999 evaded, with more female students completing them (66%) than male (59 %) (Ministry of Education, 2009).

As for social support, Lassible and Gómez (2008) realized that students who receive financial and social assistance are less likely to drop out at any time than those who do not. This issue is corroborated by Li and Chagas (2017), arguing that access to social supports is very

relevant to reduce students' dropout rates (possibly reaching a reduction of up to 7.22pp). Linked to this, the authors also show that if a student has studied in public schools during high school, their chances of dropping out in higher education reduce.

In the Brazilian context, regarding the form of admission, Li and Chagas (2017) found that the use of ENEM in the selection process increased the likelihood of dropout by 3.61%, while admission through entrance exams and new places presented a reduction of 3.72%. Also, according to Machado and Szerman (2015), the adoption of ENEM as a form of entry increases the chances of a first-year student to evade by 4.4%.

The literature also points out to a positive relationship between participation in extracurricular activities (such as research and extension) and graduation rates. This is in line with the idea defended by Tinto (2000, 2002 and 2003) regarding processes of social and academic integration having a positive impact on evasion.

Li and Chagas (2017) also found a relationship between dropout and racial aspects; according to their studies, black students drop out less than white ones in the first year of graduation. In addition, Cardoso (2008) also noticed that indigenous and black students are less likely to drop out than their white colleagues.

Finally, when Adachi (2011) analyzes Students' dropout rates at UFMG (Federal University of Minas Gerais), he noticed that students who study at night have a higher dropout rate than those who study on other shifts. The Annex 1 presents a summary table of the variables and their impacts according to the literature.

3. DATABASE, TREATMENT AND METHODOLOGY

3.1. Database

In order to answer the research problem regarding the main factors that affect dropout rates in Brazil's higher education, this study will use data offered by IBGE (Brazilian Institute of Geography and Statistics), in a sample called "Higher Education Census". According to the Institution, this is the most complete statistical survey of Brazilian Higher Education Institutions and their course offerings in order to provide detailed information on trends and situations in the sector, and to guide public educational policies (Brazilian Institute of Geography and Statistics, 2018).

This information is collected by filling out questionnaires handed over to the institutions and also by importing data from the E-MEC system, which is a system that connects the government with all educational networks, them being public or private. It is worth mentioning that, with the release of the Higher Education Census, the information appears as official statistics of higher education, becoming available for consultancy and ensuring confidentiality of the participants, no longer being possible to make any changes on the data.

Regarding this database, it has 12.043.993 records in total, since it includes all students who are in Higher Education (either in Public or Federal Universities, dropout or attending students, etc.). In order to treat the dataset, SAS Studio (statistical software) was used and, for modeling, this study used the SPSS (Statistical Package for the Social Sciences), from IBM.

3.2. Treatment

Before performing any type of statistical modeling, it is important to treat the database. This includes the need to transform data, filter and create new variables, analyze outliers (which can skew the created model) and take care of the missing values. Therefore, the following treatments were performed:

First, evasion will be analyzed in the context of Private HEIs and in undergraduate courses; thus, these were the first filters applied. In addition, only students with declared race as white, black or indigenous were considered. Moreover, students that were not studying in the night shift or full time were discarded. As for the blank values, the variables STUDENT_FINANCING and PUBLIC_HIGH_SCHOOL had some missing values which were removed.

Another important treatment concerns outliers. In the database, the only variable that is not binary is AGE. Thus, as a cause of knowledge, records whose age were less than 17 years were removed (students under 16 years of age in Brazil cannot enter the University). On the other hand, to calculate the maximum age limit, the IQR methodology (Inter Quartile Range, which refers to the difference between the third and the first quartile of the variable value) was used. As a result, any observations with a value greater than Q3 + 1.5 x IQR were removed. In the case of AGE variable, the IQR is 10, and the third quartile is 31; consequently, any variable with a value greater than 46 has been removed.

In addition, this database has a "Class' Imbalance" problem. According to Blanchard et al (2019), most classification tasks involve a situation where a class (or multiple classes) is / are over represented. Some examples where this occurs are in fraud, spam detection and cancer detection models, etc. In this database, the number of dropout students is 831.423, while the number of non-dropout students is 2.103.542 (28% and 72% of the data set, respectively), which indicates a Class' Imbalance problem. To solve this, two models were created: one with a technique called "Random Undersampling" (where records of the majority class are eliminated randomly until they reach the same number of records in the minority class) and the other with SMOTE (Synthetic Minority Oversampling Technique). In SMOTE, a subset of the minority class is removed, and a similar replica of this subset is created, which in turn is added to the main set, until the number of records of the under-represented class equals that of the over-

represented class (Blanchard et al, 2019). As the model was better when applying the first technique, it was the one selected to continue. As a result, the database had 1.662.846 records, constituted of 831.423 records for each class (evaded and not evaded). Considering that there are 12 chosen variables to analyze, this gives around 138.571 observations per variable.

Finally, the importance of each variable was assessed using the Decision Trees methodology, which is widely used for classification tasks, since they work based on the concept of finding the target variable by learning through decision rules ("if not else "), so they can find the most relevant variables in a database. Thus, it was noted that the variables with biggest impact in the dropout of students in Higher Education are: students' financing (H_4), extracurricular activities, age, being in the first year of graduation (H_1) and distance learning (H_2). This importance was also raised in order to assess if this research's hypothesis were relevant for statistical modelling. Annex 2 shows the importance of each variable in the model.

3.3. Methodology

For this article, a quantitative research design will be used. One of the reasons why this methodology was chosen is because in the available data it is possible to perceive different attributes of each student that have the potential to affect their retention in Higher Education. Thus, one of the best ways to analyze the impact of each one of these variables is through a multivariate statistical analysis, allowing to see how much each factor relates to each student's dropout in higher education. Another reason why this design was selected is that it has a particularity compared to others: estimating the magnitude of the effects of interaction between variables.

Furthermore, this study has a deductive nature, since it is based on previously developed theories that served as a basis for the creation of hypotheses, which will be proved (or disapproved) in the empirical study (Littoseliti, 2018). It will also be a cross-sectional analysis,

which means that a large amount of data was collected at a point in time, obtaining a view of how the scenario was at that moment; and it will have a conclusive character, since the objective is to test hypotheses and relationships with a very large sample of data, the data analysis will be quantitative, the information is very well defined and the evidence found can be used to support decision making by Private Higher Education Institutions (Nargundkar, 2008).

Within the types of quantitative research, there are several ways to do an analysis. For this study, Logistic Regression will be used. This modeling was chosen because it is useful to predict the presence or absence of a characteristic or result based on the values of a set of independent and predictive variables. In addition, the dependent variable is binary, representing the probability that some students will evade or not, being represented by 1 or 0; therefore, this methodology is well applicable to this research problem. This technique also aims to separate two groups of observations within a database, in order to calculate the likelihood that an observation will belong to one of these groups previously determined, which is based on the behavior of the explanatory variables (Hair, Anderson, Tatham & Black, 1998). This model is given by:

$$\textbf{Probability (Y = 1)} = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k}}$$

Where Y = 1 if the student dropped out of higher education and Y = 0 if the student did not drop out. X_1, X_2, ..., X_k are the independent variables while β_1, β_2, ..., β_k are the parameters (that is, the model coefficients for each of the explanatory variables). Also, this model is represented by a sigmoid function: when imagining a variable with values generated between -10 and 10, a sigmoid function (characteristic of Logistic Regression) defines the output between 0 and 1, expecting a relationship where the probability tends to 1 when it increases, and tends

to 0 when it decreases. Annex 3 shows the expected relationship between an explanatory variable (X) and the sigmoid curve.

4. DATA ANALYSIS

4.1. Descriptive Analysis

The average age of students in higher education at private institutions in Brazil is 26.6 years; this number makes sense because, generally, younger students end up entering Federal Universities. In addition, as inequality in the country is large, many students from Private Institutions must work and earn an income in order to be able to pay for a college. First year students represent 26.6% of the sample. The Distance Learning modality participates in the database with 28.4%; although tuition fees for distance studies are considerably lower than for face-to-face studies (in addition, this modality allows flexibility in time, and is therefore a good choice for housewives, mothers or people who are always traveling), some people still have a certain prejudice against it, which explains why there are about ¼ of students enrolled in this modality. It can also be seen that there are not many students in the disciplines of Science, Technology, Engineering and Mathematics (19.2% of the base). This also makes sense because these disciplines are more difficulty than others, and as Public Education in Brazil is not very good (and the majority of the population comes from lower economic classes), people do not feel they are well prepared to take these courses.

As for students' financing, 35.5% of students are aided by it, making it easier to pay for their education; since courses are normally expensive, and the majority of the population does not have large incomes, it makes sense that government aid (mainly through student finance) is important to help them. Also, 56.3% of the students are female, and from the whole universe of students enrolled in private HEIs, 72.1% are from public high schools; when considering that basic and secondary public education is much worse than in private schools, it makes sense that

most students in the public sector enter Private Universities; this also happens because selection processes for Federal and Public Universities are way more difficult and, consequently, these opportunities generally go to students who attended private schools and are, therefore, better prepared.

As for the form of admission, 12.7% of the sample entered Private Institutions through ENEM; these numbers make sense since ENEM works as one of the ways that a person can take to enroll in both Federal (Public) and Private Universities; despite that, the biggest part of people who participate in this process do not have in mind to enter Private Universities (they prefer to attempt a Public University's opportunity); this explains why the majority of students entering Private Higher Education go through traditional entrance exams. It is also noted that 15.1% of students participate in extracurricular activities, 12.8% are Black and 1.4% are Indians. As for the period, approximately 65.6% of the base studies at night; as most private university students must work to pay tuition, it makes sense that over a half of them study at this shift.

By crossing these variables with evasion, some interesting insights can be seen. It is noticed that students in the first year drop out less than in all other years altogether (43.1% dropped out); as for the modality, 54.9% of Distance Learning students dropped out, and 51.2% of those enrolled in CTEM subjects dropped out. Of those receiving student funding, 67.2% do not drop out, and of all women 52% continue their studies. Also, there isn't a big difference in retention rates between those who came from public schools and those who didn't (50.6% dropped out and 49.4% are still studying). As per the form of admission, of the students who used ENEM 56.6% still attend classes and, of those who take part in extra-curricular activities, only 27.9% dropped out. Regarding race, 52% of all Black students in the sample evaded, and 55% of Indian students also did it. Finally, about the shift, the number of drop-out students at night is 49.8%. These data are illustrated in Annex 4.

4.2. Modelling

Before adjusting the model, multicollinearity was verified between the variables; however, it was noted that there was no high correlation between the covariants, allowing to progress through the modelling. In addition, 6 different models were created: 1 for each hypothesis (totaling 4) and the control variables, 1 derived from the results of H_1, and 1 with all the variables to understand the impact they have on evasion.

The first model was to test the first hypothesis, that **students in the first year of Higher Education are more likely to evade.** Among the control variables, age, attending a public high school, being Black or Indian and studying at night increase chances of dropping out in Higher Education. On the other hand, being female, having enrolled through ENEM and participating in extracurricular activities reduce these chances. With the exception of race (black and indigenous), having attending a public school during high school and having joined through ENEM, the results are in accordance with the literature. Finally, results regarding being a freshman reduces chances of dropout by 33%, and this isn't in accordance with research on the subject; therefore, **we reject H_1**. Annex 5 presents the results of this model.

Since the result of this hypothesis testing was contrary to the literature, and one of the contributions of this work is to improve it, another model was created where now first year was thrown into the constant, and dummies where created for second, third, fourth, fifth and more than five years. Hence, the model shows that indeed chances of evading are higher for any year besides the first one, whereas students with more than 5 years in Universities have 95.1% more chances of evading, followed by second years and third years, whose chances are raised by 76.5% and 47.4%, respectively. Although other years beside these also raise drop-out chances, they are not that significant. As for the other variables, results are consistent with the first model. Please refer to Annex 6 in order to see all of the results regarding this model.

The third model was created with the purpose of testing whether **EAD students are more likely to drop out than in classroom teaching.** It was noticed that age, having studied in a public school, not being White, and studying at night impact positively in chances of evading (however, one must be careful with this modeling because those who study in a distance learning modality don't have any relation with the course shift, because in this case students make their schedules, and since we are isolating other variables to test the Hypothesis, there is a great chance of this specific model being biased). On the other hand, the fact of being a woman, having joined through ENEM and practicing extracurricular activities significantly reduces the chance of evasion. In this model, race, having attended a public high school and having entered through ENEM are not in accordance with the literature. As for the teaching modality, being in distance learning courses significantly increases the chances of dropping out by 126%; thus, **we accept H_2**. This result is in line with other studies related to school dropout. Annex 7 shows the results of this model.

In order to analyze the impact of courses in the area of STEM (Science, Technology, Engineering and Mathematics) on evasion, a fourth model was created. On it, race, having studied at a public school, and enrolling through ENEM are contrary to the literature, although the other factors are in accordance with it. Still regarding the variables, results are in accordance with the first two models. Thus, being in a STEM course increases the chances of evasion by 1%, making **H_3 an accepted hypothesis** and corroborating with the literature. These results can be seen in Annex 8.

The fifth model shows that **Student Financing contributes significantly to improve retention rates**, as having access to it **reduces the chances of dropout by 67%**, thus **proving H_4**. The impact of control variables is in line with previous modelling. With the exception of race, having studied in a public school and having joined through ENEM, these results are in accordance with the research. More details can be seen in Annex 9.

Finally, a last model was created containing all the explanatory variables of the database. In the analysis, results from all the previously hypotheses testing was confirmed (being in the first year and having access to student financing reduces the chances of dropout by 32% and 67%, as well as being from the distance learning modality and enrolling in Science, Technology, Engineering and Mathematics courses increase these chances by 145% and 6%, respectively). There are also some interesting points confirmed by the literature related to the covariates: with each year older, the chances of evading increase by 1%; being a woman decreases these chances by 16%; those who participate in extracurricular activities are 65% less likely to give up their degree; indigenous people are 5% less likely to evade than white, and studying at night increases these chances by 125%. Contrary to the literature, studying in public schools during high school increases the chances of dropout by 9%; having joined through ENEM reduces them by 8%, and being Black increases chances of dropping out by 14% when compared to being White. In general, this work converges with the international point of view. The model with all variables can be seen in Annex 10.

5. Conclusions and Limitations of Research

When analyzing results, there is a possibility of predicting the expected behavior of students in Private Higher Education in order to make decisions aimed at improving retention. Thus, with this information, private HEIs can direct their policies in order to earn more profit, and to contribute to a country's development by making decisions based on the expected behavior of each type of student.

As seen, students with funding and that practice extracurricular activities are the best to leverage the lifetime value of a Private Higher Education Institution. Hence, some interesting policies that a Private University could create are: Private Financing (by the institutions themselves) to acquire more students and gain greater value in the long run; partnerships with

private companies with the purpose of helping students to engage in more extracurricular activities; and creating new activities (or emphasizing the importance of existing ones), consequently increasing retention's level (some examples of extracurricular activities could be clubs ranging from practice of sports and competitions, music, cinema and arts and the creation of junior companies for students who may not be able to enter any internship). It is important to note that this is already a consolidated strategy, mainly in Asian countries, which as of today are considered references in development and study.

On the other hand, students of distance learning, attending night classes, studying Science, Technology, Engineering and Mathematics and that attended public high schools are more likely to drop out. Hence, Institutions could create and transform distance learning courses into semi-presential, because it is possible to combine the best of both worlds (flexibility and practical study), lessening the prejudice that some people may have against this type of education. That being done, Universities should start a movement to transfer students taking classes at night to these semi-presential modalities, which could result in higher retention rates. This is a possible solution because, in many cases, students who attend night classes work during the day and sometimes are too tired or have some kind of problem which they must solve, leading them to skip classes (which in the long run could result in dropout); thus, having more flexibility could help them. In addition, regarding STEM subjects and the quality of schools, since a majority of students enrolled in private HEIs come from public schools, they do not have a solid basis for more complex subjects (both intellectually and practically); therefore, the creation of workshops and monitoring for students with greater difficulty could reduce their concerns regarding the subjects and the difficulty of following classes (a problem that, if not solved, could lead to evasion).

Recommendations for future research should be considered. Firstly, information about access to transportation, notes on subjects, reasons for dropout, among other particular

information could help to better fit the profile of students more likely to evade. Secondly, it would be interesting to collect data from a private HEI to test whether these "macro" factors have the same impact at a more detailed level through case studies. Moreover, this study could be complemented with a qualitative research, since quantitative analysis doesn't allow to go into more details about internal and implicit factors that can lead to evasion. Finally, it's important to note that this model could have potential endogeneity's problems, since these results are merely associations, and shouldn't be seen as causal relations; hence, a way to infer causality should be pursued; adopting one of the strategies proposed and testing it with a treatment group, for later comparison with a holdout sample could be a suggestion.

As main contributions, this work reinforced the national literature on evasion, since despite having some studies related to this subject in the country, there is still room for improvement; also, more systematic and complex analysis on the subject within the country is welcomed, because this is still an incipient subject and with a great growth potential (Li and Chagas, 2017). Moreover, the article provides evidence that is consistent with previous studies and with the international literature, which reinforces and contributes to the academia. Likewise, it offers empirical evidence that can be used by managers of private HEIs to guide their decision-making processes. Additionally, evasion was also analyzed under a more diverse and inclusive lens (reflected in the different variables that were taken into account), because as already highlighted by Hewitt and Rose-Adams (2013), a big part of the literature on students' retention focus more on traditional factors pertaining the students (younger people, not "leaders" of their households, having a good income or not, if they work or not, among others) and hardly take into account other important points such as teaching modality, forms of entry, and so on. Finally, this study helped to improve research on tertiary education's evasion in the context of Latin America (composed mainly of developing countries), since there isn't much research revolving around this subject in the region (Venegas-Muggli, 2019).

6. BIBLIOGRAPHY

1. Adachi, Ana Amélia Chaves Teixeira. 2009. “Evasão e evadidos nos cursos de graduação da Universidade Federal de Minas Gerais”. 214f. Dissertação (Mestrado em Educação) – Programa de Pós-Graduação em Educação, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte.

2. Allmendinger J. 2019. “Adult Education and Lifelong Learning.” In Education as a Lifelong Process, edited by Hans-Peter Blossfeld and Hans-Günther Roßbach, 325-346. Wiesbaden: Edition ZfE, vol 3. Springer.

3. American College Testing Program. 1998. National college dropout and graduation rates (Iowa City, IA, American College Testing Program).

4. Andriola, Wagner Bandeira; Andriola, Cristiany Gomes; Moura, Cristiane Pascoal. 2006. “Opiniões de docentes e de coordenadores acerca do fenômeno da evasão discente dos cursos de graduação da Universidade Federal do Ceará (UFC)”. Ensaio: Avaliação e Políticas Públicas em Educação 14 (52): 365-382.

5. Archer, L. 2000. “Social Class and Access to Higher Education: A Report of findings from the Social Class and Widening Participation to HE Project”. University of North London.

6. Arulampalam, W., R. Naylor, and J. Smith. 2004. Factors affecting the probability of first year medical student dropout in the UK: A logical analysis for intake cohorts of 1980–92. Medical Education 38, no. 5: 492–503.

7. Astin, A. & Oseguera, L. 2002. Degree attainment at American colleges and universities (Los Angeles, University of California, Higher Education Research Institute).

8. BERS, T. & SMITH, K., 1991, 'Persistence of community college students: the influence of student intent and academic and social integration', Research in Higher Education, 32, pp. 539-55

9. Betsy O. Barefoot. 2004. Higher education's revolving door: confronting the problem of student drop out in US colleges and universities, Open Learning: The Journal of Open, Distance and e-Learning, 9-18.

10. Boeren, E. 2017. “Understanding Adult Lifelong Learning Participation as a Layered Problem.” Studies in Continuing Education 39 (2): 161–175

11. Boeren, E., I. Nicaise, and H. Baert. 2010. “Theoretical Models of Participation in Adult Education: The Need for an Integrated Model.” International Journal of Lifelong Education 29 (1): 45–61.

12. C. Venuleo, P. Mossi & S. Salvatore. 2016. Educational subculture and dropping out in higher education: a longitudinal case study, Studies in Higher Education, 321-342.

13. CABRERA, A., CASTANEDA, M., NORA, A. & HENGSTLER, D., 1992, "The convergence between two theories of college persistence', Journal of Higher Education, 63, pp. 143-6

14. CARDOSO, C. B. 2008. Efeitos da política de cotas na Universidade de Brasília: uma análise do rendimento e da evasão. 2008. 134 f. Dissertação (Mestrado em Educação) – Universidade de Brasília, Brasília, DF.

15. Carr, S. 2000. As distance education comes of age, the challenge is keeping the students, The Chronicle of Higher Education, A39. Available online at: http://.chronicle.com/ prm/weekly/v46/i23/23a00101.htm (Accessed 01 October 2019).

16. Cathy Schofield & Harriet Dismore. 2010. Predictors of retention and achievement of higher education students within a further education context, Journal of Further and Higher Education, 207-221.

17. Cerdán-Infantes, P., & Blom, A. 2007. Colombia: Assisting students from low-income families to attend tertiary education. En Breve, (100) The World Bank.

18. Christopher K. Morgan & Maureen Tam. 1999. Unravelling the complexities of distance education student attrition, Distance Education, 96-108.

19. Colin Beer & Celeste Lawson. 2017. The problem of student attrition in higher education: An alternative perspective, Journal of Further and Higher Education, 773-784.

20. Colin Beer & Celeste Lawson. 2018. Framing attrition in higher education: a complex problem, Journal of Further and Higher Education, 497-508.

21. Colin Beer & Celeste Lawson. 2018. Framing attrition in higher education: a complex problem, Journal of Further and Higher Education, 497-508.

22. COSTA, C. J.; COCHI, C. B. R. 2014. A expansão do Ensino Superior no Brasil e a Educação a Distância: instituições públicas e privadas. Teoria e Prática da Educação, v. 16, n. 1, p. 21-32.

23. Deming, David J., Claudia Goldin, Lawrence F. Katz, and Noam Yuchtman. 2015. Can online learning bend the higher education cost curve? American Economic Review 105, no. 5:496–501.

24. Department of Higher Education and Training (DHET). 2016. First Time Entering Undergraduate Cohort Studies for Public Higher Education Institutions. Pretoria: Department of Higher Education and Training, South Africa.

25. DesJardins, S. J., Ahlburg, D., & McCall, B. P. 2006. An integrated model of application, admission, enrollment, and financial aid. The Journal of Higher Education, 77(3), 381–429.

26. DEST Strategic Analysis and Evaluation Group. 2004. Higher education attrition rates 1994 – 2002: A brief overview. Retrieved 12 February, 2007, from http://www.dest.gov.au/sectors/higher_education/publications_resources/statistics/higher_education_attrition_rates_ 1994_2002.htm

27. DfEE. 1998. Participation in Education and Training by 16–18 years old in England:1987 to 1997. Statistical Bulletin 335/ 98.

28. Diamond, A. & Sekhon, J. 2013. Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies. The Review of Economics and Statistics. July 2013, Vol. 95, No. 3, Pages: 932-945.

29. DIAS, Ellen C. M.; THEÓPHILO, Carlos R.; LOPES, Maria A. S. 2010. Evasão no Ensino Superior: estudo dos fatores causadores da evasão no curso de Ciências Contábeis da Universidade Estadual de Montes Claros - Unimontes - MG. In: CONGRESSO USP DE INICIAÇÃO CIENTÍFICA EM CONTABILIDADE.

30. Dynarski, S. 2002. The behavioral and distributional implications of aid for college. American Economic Review, 92(2), 279–285.

31. Fregoneis, Jucelia Geni Pereira. 2002. "Estudo do Desempenho Acadêmico nos Cursos de Graduação dos Centros de Ciências Exatas e de Tecnologia da Universidade Estadual de Maringá: período 1995 – 2000". Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina (UFSC), Florianópolis.

32. Gérard Lassibille & Lucía Navarro Gómez. 2008. Why do higher education students drop out? Evidence from Spain, Education Economics, 89-105.

33. Govendir, M., Ginns, P., Symons, R., & Tammen, I. 2009. Improving the research higher degree experience at the Faculty of Veterinary Science, The University of Sydney, in The

student experience, Proceedings of the 32nd HERDSA Annual Conference, Darwin, pp. 163–172.

34. GROSSET, J., 1991, 'Patterns of integration, commitment, and student characteristics, and retention among younger and older students', Research in Higher Education, 32, pp. 159-78.

35. HE, W.; XU, G.; KRUCK, S. E. 2014. Online IS education for the 21st century. Journal of Information Systems Education, v. 25, n. 2, p. 101.

36. Heenan, D. 2002. "Women, Access and Progression: an Examination of Women's Reasons for not Continuing in Higher Education Following the Completion of the Certificate in Women's Studies." Studies in Continuing Education 24 (1): 39–55.

37. Heublein, U. 2014. "Student Drop-out from German Higher Education Institutions." European Journal of Education 49 (4): 497–513. doi:10.1111/ejed.12097.

38. Heublein, U., H. Spangenberg, and D. Sommer. 2003. Ursachen des Studienabbruchs. Analyse 2002. Hanover: Hochschul Informations System.

39. Heublein, U., H. Spangenberg, and D. Sommer. 2003. Ursachen des Studienabbruchs. Analyse 2002. Hanover: Hochschul Informations System.

40. Heublein, U., Schmelzer, R., & Sommer, D. 2008. Die Entwicklung der Studienabbruchquote an den deutschen Hochschulen. Ergebnisse einer Berechnung des

Studienabbruchs af der Basis des Absolventenjahgangs 2006. [The development of student drop-out in the German higher education system. Results from an analysis of the non-completion based on the cohort of 2006.] HIS:Projektbericht. Hannover: HIS (Hochschul-InformationsSystem GmbH). (http://www.bildungsserver.de/zeigen.html?seite=4605) (Accessed 01 October 2019).

41. Hewitt, L., and J. Rose-Adams. 2013. "What'retention'means to me: The Position of the Adult Learner in Student Retention." Widening Participation and Lifelong Learning 14 (1): 146–164.

42. Ho et. Al. 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Political Analysis, 15(3).

43. HOED, R. M. 2016. Análise da evasão em cursos superiores: o caso da evasão em cursos superiores da área de Computação. 2016. 188 f. Dissertação (Mestrado Profissional em Computação Aplicada) – Universidade de Brasília, Brasília.

44. Holmegaard, H. T., L. M. Madsen, and L. Ulriksen. 2014. "A Journey of Negotiation and Belonging: Understanding Students' Transition to Science and Engineering in Higher Education." Cultural Studies of Science Education 9 (3): 755–786.

45. INEP: Censo da Educação Superior. 2013. Available in: <http://portal.inep.gov.br/basica-levantamentos-microdados>. Accessed on: 01 October 2019.

46. INEP: Censo da Educação Superior. 2014. Available in: <http://portal.inep.gov.br/basica-levantamentos-microdados>. Accessed on: 01 October 2019.

47. INEP: Censo da Educação Superior. 2016. Available in: <http://portal.inep.gov.br/basica-levantamentos-microdados>. Accessed in: 01 October 2019.

48. INEP: Censo da Educação Superior. 2018. Available in: <http://portal.inep.gov.br/basica-levantamentos-microdados>. Accessed in: 28 February 2020.

49. Jeffreys, M.R. 2007. Tracking students through program entry, progression, graduation and licensure: Assessing undergraduate nursing student retention and success. Nurse Education Today 27, no. 5: 406–419.

50. Jenny Roberts. 2018. Professional staff contributions to student retention and success in higher education, Journal of Higher Education Policy and Management, 140-153.

51. Joshua Goodman & Julia Melkers & Amanda Pallais, 2019. "Can Online Delivery Increase Access to Education?" Journal of Labor Economics, vol 37(1), pages 1-34.

52. Juan I. Venegas-Muggli. 2019. Higher education dropout of non-traditional mature freshmen: the role of sociodemographic characteristics, Studies in Continuing Education.

53. Kember, D. 1995. Open learning courses for adults: A model of student progress. Englewood Cliffs, NJ: Educational Technology.

54. Krause, K.-L., Hartley, R., James, R., & McInnis, C. 2005. The first year experience in Australian universities: Findings from a decade of national studies. Canberra: Centre for the

Study of Higher Education, University of Melbourne and Department of Education, Science and Training.

55. Lee, S. & Pradhan, B. 2007. “Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models.” In: “Landslides”, p. 4 to 33.

56. LI, D. L.; CHAGAS, A. L. S. 2017. Efeitos do Sisu sobre a migração e a evasão estudantil.

57. Lima Jr., Paulo; Silveira, Fernando Lang da; Ostermann, Fernanda. 2002. “Análise de Sobrevivência aplicada ao estudo do fluxo escolar nos cursos e graduação em física: um exemplo de uma universidade brasileira”. Revista Brasileira de Ensino em Física 34 (1): 1-10.

58. Litosseliti, L. 2018. “Research Methods in Linguistics”. 2nd ed. United Kingdon: Bloomsbury Publishing.

59. LITTO, F. M.; FORMIGA, M. 2008. Educação a distância: o estado da arte. São Paulo: Pearson.

60. Liz Thomas. 2002. Student retention in higher education: the role of institutional habitus, Journal of Education Policy, 423-442.

61. Losty, B. P. & Broderson, D.D. 1980, 'Who succeeds? A comparison of transcripts of graduates and inactive students of a nontraditional bachelor of arts program', Alternative Higher Education: The Journal of Nontraditional Studies, no. 25, pp. 91-9.

62. MACHADO, C.; SZERMAN, C. 2015. The effects of a centralized college admission mechanism on migration and college enrollment: Evidence from Brazil. Working Paper LACEA.

63. Mark Pearson. 2012. Building bridges: higher degree student retention and counselling support, Journal of Higher Education Policy and Management, 187-199.

64. Martin, Y.M., Maclachlan, M., & Karmel, T. 2001. Postgraduate completion rates. Canberra: Commonwealth of Australia.

65. McInnes, C., James, R., & Hartley, R. 2000. Trends in the first year experience in Australian universities, 2000. Melbourne: Department of Employment, Education, Training and Youth Affairs.

66. MEC/ANDIFES/ABRUEM/SESU. 1996. Diplomação, retenção e evasão nos cursos de graduação em instituições de Ensino Superior públicas. Avaliação, Campinas, SP, v., n. 2, p. 55-65, 1996.

67. Melguizo, T. 2011. A review of the theories developed to describe the process of college persistence and attainment. In J. C. Smart (Ed.), Higher education: Handbook of theory and research.

68. Ministerio de Educación y Ciencia. 2001. Informe transversal del rendimiento académico de las ingenierías técnicas. Madrid: Consejo de Universidad.

69. Ministry for Higher Education and Research. 2016. Repères et Références Statistiques [Benchmarks and Statistical References]. Paris: DEPP.

70. Ministry of Education. 2009. Draft tertiary education strategy, 2010–2015. Wellington: Ministry of Education.

71. Nargundkar, R. 2008. "Marketing Research: Text and Cases", Tata McGraw-Hill Educational, p.39.

72. National Audit Office (NAO). 2002. Improving student achievement in English higher education. Report by the Comptroller and Auditor General (London: The Stationary Office).

73. National Audit Office (NAO). 2007. Staying the course: The retention of students in higher education. London: Stationery Office.

74. National Audit Office. 2007. Staying the course: The retention of students in higher education. London: The Stationery Office.

75. NATIONAL CENTER FOR EDUCATION STATISTICS (NCES). 1995. The Condition of Education, 1995 (Washington, DC, US Department of Education).

76. NATIONAL CENTER FOR EDUCATION STATISTICS (NCES). 1996. The Condition of Education 1996 (Washington, DC, US Department of Education).

77. Nora, Amaury. 1990. “Campus-based Aid Programs as Determinants of Retention among Hispanic Community College Students”. The Journal of Higher Education 61 (3): 312-331.

78. OECD. 2018. Education at Glance 2018: OECD Indicators, OECD Publishing, Paris, http://dx.doi.org/10.1787/eag-2018-en.

79. OECD: Education at a Glance 2015: OECD indicators 2015, OECD 2015, (Brasil—Country Note), pp. 1–10.

80. OECD: Education at a glance 2016: OECD indicators, pp. 507 (2016).

81. PASCARELLA, E & CHAPMAN, D., 1983, 'A multi-institutional, path analytic validation of Tinto's model of college withdrawal', American Educational Research Journal, 20, pp. 87-102.

82. PASCARELLA, E. & TERENZINI, P., 1980, 'Predicting freshman persistence and voluntary dropout decisions from a theoretical model', Journal of Higher Education, 51, pp. 60-75.

83. PETERSON, S. & DELMAS, R. 1996. 'Effects of career decision-making self-efficacy on the retention of underprepared adults: a path-analytic study of student persistence', paper presented to the American Education Research Association, New York, April 1996.

84. Porter, M. E. 1985. The Competitive Advantage: Creating and Sustaining Superior Performance. NY: Free Press.

85. Pritchard, S. 1995. A higher degree of loneliness. The Independent Education News, Retrieved 30 August, 2011, from http://www.independent.co.uk/news/education/educationnews/a-higherdegree-of-loneliness-1584514.html

86. Reay, D. 2002. “Class, Authenticity and the Transition to Higher Education for Mature Students.” The Sociological Review 50 (3): 398–418.

87. Romuald Bodin & Sophie Orange. 2018. Access and retention in French higher education: student drop-out as a form of regulation, British Journal of Sociology of Education, 126-143.

88. Rosenbaum, P. & Rubin, D. 1983. The central role of the prop. score in observational studies for causal effects. Biometricka. Apr. 1.

89. ROSENBAUM, P. R. 2002. Observational Studies, 2nd ed. Springer, New York.

90. Saccaro, França & Jacinto. 2019. Estud. Econ., São Paulo, vol.49 n.2, p.337-373.

91. Selden, W.K. 1965. The governance of higher education.

92. Seung Youn Chyung. 2001. Systematic and systemic approaches to reducing attrition rates in online higher education, American Journal of Distance Education, 36-49.

93. Shari L. Peterson, Patricia Kovel-Jarboe & Sherry A. Schwartz. 1997. Quality Improvement in Higher Education: implications for student retention, Quality in Higher Education, 131-141.

94. Sheppard, R. S. 2002. Predictors of student success in distance education courses. Unpublished doctoral dissertation, West Virginia University, Morganstown, West Virginia.

95. Shrivastava, M., Shrivastava, S. 2014. Political economy of higher education: comparing South Africa to trends in the world. High. Educ. 67, 809–822.

96. SILVA, F. I. C. et al. 2012. Evasão escolar no curso de Educação Física da Universidade Federal do Piauí. Avaliação, Campinas, SP, v. 17, n. 2, p. 391 – 404.

97. Spady, W. G. 1970. "Dropouts from Higher Education: An Interdisciplinary Review and Synthesis." Interchange 1: 64–85.

98. Swail, W. S., Redd, K. E., & Perna, L. W. 2003. Retaining minority students in higher education: A framework for success. San Fransico, CA: Jossey-Bass.

99. Teichler, U. 2014. Opportunities and problems of comparative higher education research: the daily life of research. High. Educ. 67, 393–408.

100. The Organisation for Economic Co-operation and Development (OECD). 2008. Encouraging Student Interest in Science and Technology Studies. Paris: Global Science Forum.

101. Tinto, V. 1975. “Drop-Out from Higher Education: A Theoretical Synthesis of Recent Research.” Review of Education Research 45: 89–125.

102. Tinto, V. 2000. Taking student retention seriously: Rethinking the first year of college. NASCADA conference proceedings.

103. Tinto, V. 2002. Promoting student retention: Lessons learned from the United States. Paper presented at the 11th Annual Conference of the European Access Network, June 19, in Prato, Italy.

104. Tinto, V. 2003. Promoting student retention through classroom practice. Paper presented at Conference on Enhancing Student Retention: Using International Policy and Practice, November 5–7, in Amsterdam, The Netherlands.

105. TINTO, V., 1975, 'Dropout from higher education: a theoretical synthesis of recent research', Review of Educational Research, 45, pp. 89-125.

106. TINTO, V., 1987, Leaving College: Rethinking the causes and cures of student attrition (Chicago, University of Chicago Press).

107. TINTO, V., 1993, Leaving College: Rethinking the causes and cures of student attrition, 2nd edn (Chicago, University of Chicago Press).

108. Tinto, V., and A. Goodsell. 1993. A Longitudinal Study of Freshman Interest Groups at the University of Washington. University Park, PA: National Centre for Postsecondary Teaching, Learning, and Assessment.

109. VANZ, S. A. S. et al. 2016. Evasão e retenção no curso de Biblioteconomia da UFRGS. Avaliação, Campinas; Sorocaba, SP, v. 21, n. 2, p. 541 – 568.

110. Werner Georg. 2009. Individual and institutional factors in the tendency to drop out of higher education: a multilevel analysis using data from the Konstanz Student Survey, Studies in Higher Education, 647-661.

111. Wintre, M.G., Bower, C., Gordner, N., & Lange, L. 2006. Re-evaluating the university attrition statistic: A longitudinal follow-up study. Journal of Adolescent Research, 21(2), 111–132.

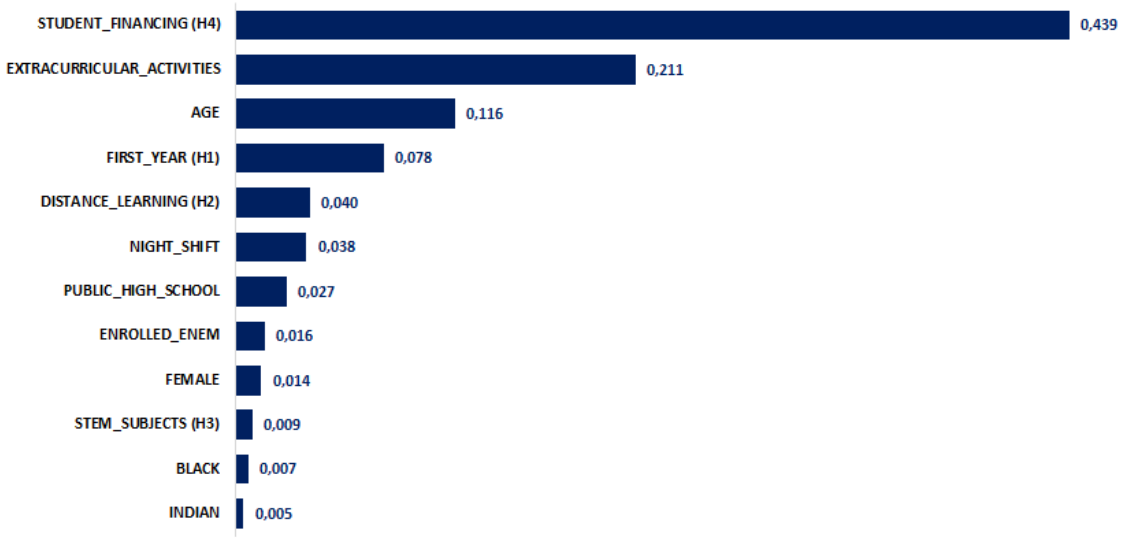
112. Woosley, S. A. 2003. "How Important are the First Few Weeks of College? The Long-Term Effects of Initial College Experiences." College Student Journal 37 (2): 201–207.

7. ANNEXES

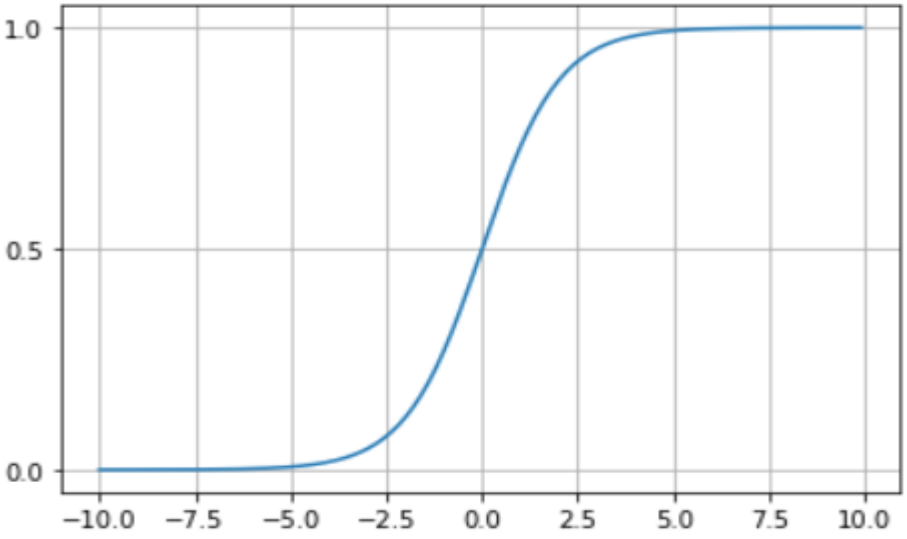
Annex 1: Table of variables and its impact on Evasion

Variables	Description	Impact on Evasion	References
FIRST_YEAR (H_1)	1. Student is a freshman 0. Student isn't a freshman	Students on the first year of Universities have more chances of evading that on other years	- Venegas-Muggli (2019); - Barefoot (2004); - Bodin e Orange (2018); - Lassible e Gómez (2008); - Kember (1995)
DISTANCE_LEARNING (H_2)	1. Student enrolled in Distance Learning 0. Student not enrolled in Distance Learning	Students taking Distance Learning courses are more susceptible to evade than on traditional face to face classes	- Carr (2000); - Losty e Broderson (1980); - Chyung (2001); - Swail, Redd e Perna (2003); - Sheppard (2002)
STEM_SUBJECTS (H_3)	1. Student enrolled in STEM Subjects 0. Student not enrolled in Distance Learning	Students taking courses in STEM (Science, Technology, Engineer and Mathematics) have more chances of dropping out than in other courses	- Heublein (2014); - OECD (2008); - Heublein, Schmelzer e Sommer (2008)
STUDENT_FINANCING (H_4)	1. Student has access to financing 2. Student doesn't have access to financing	Having access to forms of financial aid for students greatly reduces odds of evading	- DesJardins et al. (2006); - Dynarski (2002); - Melguizo (2010); - Cerdán-Infantes and Blom (2007); - Saccaro, França e Jacinto (2019); - Nora (1990); - Santelices (2016)
AGE	Numerical	When students are older, they have greater chances of evading	- Arulampalam, Naylor, and Smith (2004); - Jeffreys (2007) - Venuleo e Salvatore (2016); - Lassible e Gómez (2008)
FEMALE	1. Student is Female 0. Student is Male	Female students have less chances of dropping out than male students.	- Arulampalam, Naylor and Smith (2004); - Lassible e Gómez (2008); - NAO (2007)
PUBLIC_HIGH_SCHOOL	1. Student attended public high school 0. Student didn't attend public high school	Students who attended public high schools are more likely to give up their subjects than students who attended private schools.	- Li e Chagas (2017)
ENROLLED_ENEM	1. Student enrolled through ENEM 0. Student enrolled through Selective Proccess	Students enrolled through ENEM have greater chances of evading than when enrolled through traditional selection.	- Li e Chagas (2017); - Machado e Szerman (2015)
EXTRACURRICULAR_ACTIVITIES	1. Student participates in Extracurricular Activities 0. Student doesn't participate in Extracurricular Activities	Extracurricular Activities have a significant positive impact on retention rates.	- Tinto (2000, 2002 e 2003)
BLACK	1. Student is Black 0. Student isn't Black	Black students are less likely to evade than White students	- Li e Chagas (2007); - Cardoso (2008)
INDIAN	1. Student is Indian 0. Student isn't Indian	Indian students are less likely to evade than White Students	- Li e Chagas (2007); - Cardoso (2008)
NIGHT_SHIFT	1. Student studies at night 0. Student doesn't study at night	Students who take courses at night are more likely to drop out than students taking full time courses	- Adachi, 2011

Annex 2: Feature Importance from Decision Trees



Annex 3: Expected relation between an explanatory variable and the sigmoid curve.



Annex 4: Descriptive Statistics

Numerical Variable	Status	Average	Std. Dev.	Min.	Max.
AGE	Total	26,6	6,8	17	46
	Evaded	27,7	6,8	17	46
	Not Evaded	26,5	6,9	17	46

Binary Variable	Status	% Share on Data	% of Variable
FIRST_YEAR	Total	26,6%	100,0%
	Evaded	23,0%	43,1%
	Not Evaded	30,3%	56,9%
DISTANCE_LEARNING	Total	28,4%	100,0%
	Evaded	31,1%	54,9%
	Not Evaded	25,6%	45,1%
STEM_SUBJECTS	Total	19,2%	100,0%
	Evaded	19,6%	51,2%
	Not Evaded	18,7%	48,8%
STUDENT_FINANCING	Total	35,5%	100,0%
	Evaded	23,3%	32,8%
	Not Evaded	47,7%	67,2%
FEMALE	Total	56,3%	100,0%
	Evaded	54,0%	48,0%
	Not Evaded	58,5%	52,0%
PUBLIC_HIGH_SCHOOL	Total	72,1%	100,0%
	Evaded	72,9%	50,6%
	Not Evaded	71,2%	49,4%
ENROLLED_ENEM	Total	12,7%	100,0%
	Evaded	11,1%	43,4%
	Not Evaded	14,4%	56,6%
EXTRACURRICULAR_ACTIVITIES	Total	15,1%	100,0%
	Evaded	8,4%	27,9%
	Not Evaded	21,7%	72,1%
BLACK	Total	12,8%	100,0%
	Evaded	13,4%	52,0%
	Not Evaded	12,3%	48,0%
INDIANS	Total	1,4%	100,0%
	Evaded	1,5%	55,0%
	Not Evaded	1,2%	45,0%
NIGHT_SHIFT	Total	65,6%	100,0%
	Evaded	65,3%	49,8%
	Not Evaded	65,9%	50,2%

Annex 5: LOGIT results for testing H_1

Variable	Hypothesis	Coefficient	Sig.	Odds Ratio	Literature
FIRST_YEAR	H_1	-0,40	0,00	0,67	⊗
AGE		0,02	0,00	1,02	✓
FEMALE		-0,17	0,00	0,84	✓
PUBLIC_HIGH_SCHOOL		0,09	0,00	1,09	⊗
ENROLLED_ENEM		-0,20	0,00	0,81	⊗
EXTRACURRICULAR_ACTIVITIES		-1,09	0,00	0,34	✓
BLACK		0,04	0,00	1,04	⊗
INDIANS		0,08	0,00	1,08	⊗
NIGHT_SHIFT		0,01	0,01	1,01	✓
CONSTANT		**-0,20**	**0,00**	**0,82**	

Annex 6: LOGIT results for other years of graduation (Derived from H_1)

Variable	Hypothesis	Coefficient	Sig.	Odds Ratio	Literature
AGE		0,02	0,00	1,02	✓
FEMALE		-0,17	0,00	0,85	✓
PUBLIC_HIGH_SCHOOL		0,09	0,00	1,10	⊗
ENROLLED_ENEM		-0,22	0,00	0,81	⊗
EXTRACURRICULAR_ACTIVITIES		-1,08	0,00	0,34	✓
BLACK		0,05	0,00	1,05	⊗
INDIANS		0,08	0,00	1,08	⊗
NIGHT_SHIFT		0,03	0,00	1,03	✓
SECOND_YEAR		0,57	0,00	1,77	⊗
THIRD_YEAR		0,39	0,00	1,47	⊗
FOURTH_YEAR		0,13	0,00	1,14	⊗
FIFTH_YEAR		0,11	0,00	1,11	⊗
OTHER_YEARS		0,67	0,00	1,95	⊗
CONSTANT		**-0,62**	**0,00**	**0,82**	

Annex 7: LOGIT results for testing H_2

Variable	Hypothesis	Coefficient	Sig.	Odds Ratio	Literature
DISTANCE_LEARNING	H_2	0,82	0,00	2,26	✓
AGE		0,02	0,00	1,02	✓
FEMALE		-0,18	0,00	0,83	✓
PUBLIC_HIGH_SCHOOL		-0,01	0,21	1,00	✗
ENROLLED_ENEM		-0,21	0,00	0,81	✗
EXTRACURRICULAR_ACTIVITIES		-1,03	0,00	0,36	✓
BLACK		0,02	0,00	1,02	✗
INDIANS		0,08	0,00	1,08	✗
NIGHT_SHIFT		0,75	0,00	2,12	✓
CONSTANT		**-0,88**	**0,00**	**0,42**	

Annex 8: LOGIT results for testing H_3

Variable	Hypothesis	Coefficient	Sig.	Odds Ratio	Literature
STEM_SUBJECTS	H_3	0,01	0,15	1,01	✓
AGE		0,02	0,00	1,02	✓
FEMALE		-0,17	0,00	0,84	✓
PUBLIC_HIGH_SCHOOL		0,07	0,00	1,07	✗
ENROLLED_ENEM		-0,22	0,00	0,81	✗
EXTRACURRICULAR_ACTIVITIES		-1,06	0,00	0,35	✓
BLACK		0,04	0,00	1,04	✗
INDIANS		0,10	0,00	1,11	✗
NIGHT_SHIFT		0,09	0,00	1,09	✓
CONSTANT		**-0,41**	**0,00**	**0,66**	

Annex 9: LOGIT results for testing H_4

Variable	Hypothesis	Coefficient	Sig.	Odds Ratio	Literature
STUDENT_FINANCING	H_4	-1,12	0,00	0,33	✓
AGE		0,02	0,00	1,02	✓
FEMALE		-0,17	0,00	0,84	✓
PUBLIC_HIGH_SCHOOL		0,14	0,00	1,15	✗
ENROLLED_ENEM		-0,11	0,00	0,90	✗
EXTRACURRICULAR_ACTIVITIES		-1,08	0,00	0,34	✓
BLACK		0,15	0,00	1,16	✗
INDIANS		0,00	0,95	1,00	✗
NIGHT_SHIFT		0,16	0,00	1,18	✓
CONSTANT		**-0,12**	**0,00**	**0,89**	

Annex 10: LOGIT results for all Hypothesis (Full Model).

Variable	Hypothesis	Coefficient	Sig.	Odds Ratio	Literature
FIRST_YEAR	H_1	-0,39	0,00	0,68	✗
DISTANCE_LEARNING	H_2	0,90	0,00	2,45	✓
STEM_SUBJECTS	H_3	0,06	0,00	1,06	✓
STUDENT_FINANCING	H_4	-1,10	0,00	0,33	✓
AGE		0,01	0,00	1,01	✓
FEMALE		-0,17	0,00	0,84	✓
PUBLIC_HIGH_SCHOOL		0,08	0,00	1,09	✗
ENROLLED_ENEM		-0,09	0,00	0,92	✗
EXTRACURRICULAR_ACTIVITIES		-1,06	0,00	0,35	✓
BLACK		0,13	0,00	1,14	✗
INDIANS		-0,05	0,00	0,95	✓
NIGHT_SHIFT		0,80	0,00	2,24	✓
CONSTANT		**-0,44**	**0,00**	**0,65**	